

# Small-sample one-sided testing in extreme value regression models

Silvia L.P. Ferrari\* and Eliane C. Pinheiro

*Departamento de Estatística, Universidade de São Paulo, Brazil*

May 26, 2014

## Abstract

We derive adjusted signed likelihood ratio statistics for a general class of extreme value regression models. The adjustments reduce the error in the standard normal approximation to the distribution of the signed likelihood ratio statistic. We use Monte Carlo simulations to compare the finite-sample performance of the different tests. Our simulations suggest that the signed likelihood ratio test tends to be liberal when the sample size is not large, and that the adjustments are effective in shrinking the size distortion. Two real data applications are presented and discussed.

*Key words:* Extreme value regression; Gumbel distribution; Nonlinear models; Signed likelihood ratio test; Small-sample adjustments.

## 1 Introduction

Statistics of extremes are applied in a wide range of areas, such as marine, civil and telecommunications engineering, metallurgy, geology, astronomy, oceanography, meteorology, hydrology, reliability, finance, insurance, medicine and nutrition. In the last few decades, the growth in the intensity and frequency of environmental phenomena – Hurricane Katrina (2005), the Japan earthquake (2011), mudslides and floods due to intense rains in Australia (2010-2011), Pakistan (2010) and Brazil (2011), the typhoon in the Philippines (2013) – and disasters caused by man – oil leaks (2010) and financial crises (2008, 2010) – has highlighted the necessity of improving the inference accuracy of the statistics of extremes.

A classical reference regarding the statistics of extremes is the study [Gumbel \(1958\)](#), who formalized the extreme value theory presenting the only three asymptotic limiting distributions of extreme values obtained by [Fréchet \(1927\)](#) and [Fisher & Tippet \(1928\)](#), and unified

---

\*Corresponding author. Email: [silviaferrari.usp@gmail.com](mailto:silviaferrari.usp@gmail.com)

by [Gnedenko \(1943\)](#). More recent book-length references include the works of [de Haan \(1970\)](#), [Coles \(2001\)](#), [Beirlant et al. \(2004\)](#), [Kotz & Nadajarah \(2000\)](#) and, [Castillo et al. \(2005\)](#), among others.

The extreme value distribution is frequently used to model extreme events data and reliability/survival outcomes. The focus of this paper is on inference for extreme value regression models when a sample is small or of moderate size. Specifically, we deal with a general class of extreme value regression models introduced by [Barreto-Souza & Vasconcellos \(2011\)](#) in which location and dispersion parameters are linear or, possibly, non-linear functions of regressors. [Mueller & Rufibach \(2009\)](#) commented on the lack of small sample extreme value theory in the literature. Nevertheless, there are many practical applications for which data sets are not large enough to safely apply asymptotic theory. Recent contributions in small-sample inference for extreme value regression models are those made by [Barreto-Souza & Vasconcellos \(2011\)](#), who obtained bias corrections for maximum likelihood estimators, and [Ferrari & Pinheiro \(2012\)](#), who derived Skovgaard’s adjustment ([Skovgaard, 2001](#)) to the likelihood ratio statistic for testing two-sided hypotheses. [Ferrari & Pinheiro \(2012\)](#) compared the small sample accuracy of the usual likelihood ratio, Wald, score and gradient tests, and the adjusted likelihood ratio tests in extreme value regression models with linear and non-linear predictors for both location and dispersion parameters. Simulation results suggest that the type I error of the likelihood ratio and Wald tests can be markedly greater than the nominal level in small- and moderate-sized samples. The gradient test presents similar behavior, but is much less oversized. The score test is even less size distorted, can be conservative in some scenarios, and is competitive with the adjusted likelihood ratio test in some cases. Overall, the adjusted likelihood ratio test performs better than all others tests.

The present paper is focused on one-sided tests, for which the above mentioned statistics are not suitable. The signed likelihood ratio statistic is the most widely used for one-sided testing of a simple null hypothesis about a scalar parameter in the presence of nuisance parameters. This statistic has a limiting standard normal distribution, but its finite-sample law is usually unknown. It is well-known that the use of the asymptotic distribution in small samples frequently lead to oversized signed likelihood ratio tests. Considerable research has been conducted to propose adjustments to improve the accuracy of the standard normal approximation to the distribution of the signed root of the likelihood ratio statistic. In this paper, we derive five adjustments, proposed by [Barndorff-Nielsen \(1986, 1991\)](#), [DiCiccio & Martin \(1993\)](#), [Skovgaard \(1996\)](#), [Severini \(1999\)](#) and [Fraser et al. \(1999\)](#), to extreme value regression models. We compare the finite-sample performance of the signed likelihood ratio test and the adjusted signed likelihood ratio tests obtained in this work. Simulation results suggest that the adjusted statistics approximately follow a standard normal distribution with a high degree of accuracy even in small samples. Furthermore, we illustrate an application of the usual signed likelihood ratio test and its modified versions in a real data set.

The paper is organized as follows. In Section 2 we present the extreme value regression models. In Section 3, the modified signed likelihood ratio statistic proposed by Barndorff-Nielsen (1986, 1991) is presented along with approximations derived by different authors, namely DiCiccio & Martin (1993), Skovgaard (1996), Severini (1999) and Fraser et al. (1999). In Section 4 we derive adjusted signed likelihood ratio statistics in extreme value regression models. Monte Carlo simulations are presented in Section 5 to compare the performances of the tests, whereas applications to real data sets are presented in Section 6. The paper ends with conclusions in Section 7. Technical details regarding DiCiccio & Martin's approach are presented in the Appendix.

## 2 Extreme value regression models

Let  $y$  be a continuous random variable with a maximum extreme value distribution with location parameter  $\mu$ , dispersion parameter  $\sigma$ , and density function

$$f(y; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{y - \mu}{\sigma}\right) \exp\left\{-\exp\left(-\frac{y - \mu}{\sigma}\right)\right\}, \quad y \in \mathbb{R}, \quad (1)$$

where  $\mu \in \mathbb{R}$  and  $\sigma > 0$ ; we write  $y \sim EV_{max}(\mu, \sigma)$ . The mean and the variance of  $y$  are  $E(y) = \mu + \mathcal{E}\sigma$  and  $\text{var}(y) = \sigma^2\pi^2/6$ , respectively, where  $\mathcal{E}$  is the Euler constant;  $\mathcal{E} \approx 0.5772$ . This distribution is also called the Gumbel or type I extreme value distribution. Here, we refer to distribution as the maximum extreme value distribution to distinguish it from the minimum extreme value distribution, which is also often called the Gumbel or type I extreme value distribution in the statistical literature. The distribution in (1) is one of the three possible limiting laws of the standardized maximum of independent and identically distributed random variables (Gnedenko, 1943), and is frequently invoked to model extreme events; see, for instance, Castillo et al. (2005, Tables 9.16) and Coles (2001, Section 3.4.1).

The density function of the minimum extreme value distribution with location parameter  $\mu$  and dispersion parameter  $\sigma$  is given by

$$f(y; \mu, \sigma) = \frac{1}{\sigma} \exp\left(\frac{y - \mu}{\sigma}\right) \exp\left\{-\exp\left(\frac{y - \mu}{\sigma}\right)\right\}, \quad y \in \mathbb{R}, \quad (2)$$

and we write  $y \sim EV_{min}(\mu, \sigma)$ . The mean of the minimum extreme value distribution is  $\mu - \mathcal{E}\sigma$  and its variance coincides with that of the maximum extreme value distribution. A useful property is that

$$y \sim EV_{min}(\mu, \sigma) \iff -y \sim EV_{max}(-\mu, \sigma). \quad (3)$$

The distribution in (2) is the distribution of the logarithm of Weibull distributed random variables, and is often used in reliability and survival analysis to model log-lifetimes (Lawles, 2003).

Let  $y_1, \dots, y_n$  be independent random variables, where

$$y_t \sim EV_{max}(\mu_t, \sigma_t), \quad (4)$$

for  $t = 1, \dots, n$ . The maximum extreme value regression model is defined by (4) and by two systematic components given by

$$g(\mu_t) = \eta_t = \eta(x_t, \beta) \quad (5)$$

and

$$h(\sigma_t) = \delta_t = \delta(z_t, \gamma), \quad (6)$$

where  $\beta = (\beta_1, \dots, \beta_k)^\top$  and  $\gamma = (\gamma_1, \dots, \gamma_m)^\top$  are vectors of unknown regression parameters ( $\beta \in \mathbb{R}^k$  and  $\gamma \in \mathbb{R}^m$ ,  $k + m < n$ ) and  $x_t$  and  $z_t$  are observations on covariates. Here,  $\eta(\cdot, \cdot)$  and  $\delta(\cdot, \cdot)$  are continuously twice differentiable (possibly nonlinear) functions in the second argument. Finally,  $g(\cdot)$  and  $h(\cdot)$  are known, strictly monotonic and twice differentiable link functions that map  $\mathbb{R}$  and  $\mathbb{R}^+$  onto  $\mathbb{R}$ , respectively. Let  $X$  be the derivative matrix of  $\eta = (\eta_1, \dots, \eta_n)^\top$  with respect to  $\beta^\top$ , and let  $Z$  be the derivative matrix of  $\delta = (\delta_1, \dots, \delta_n)^\top$  with respect to  $\gamma^\top$ . It is assumed that  $\text{rank}(X) = k$  and  $\text{rank}(Z) = m$  for all  $\beta$  and  $\gamma$ .

Analogously, we define the minimum extreme value regression model as above, by substituting the following for (4):

$$y_t \sim EV_{min}(\mu_t, \sigma_t), \quad (7)$$

for  $t = 1, \dots, n$ . Hereinafter, we shall refer to the maximum extreme value regression model (4)–(6) only, but, as will be shown, our results can be easily adapted to the minimum extreme value regression model (5)–(7).

### 3 Adjusted signed likelihood ratio statistics

Let  $\theta = (\beta^\top, \gamma^\top)^\top$  be the unknown parameter vector that indexes the extreme value regression model (4)–(6). In what follows,  $\nu$  represents the scalar parameter of interest and  $\psi = (\psi_1, \dots, \psi_s)^\top$  is the nuisance parameter; note that  $1 + s = k + m$ . Let  $\ell(\theta)$  be the log-likelihood function and let  $I$  and  $J$  be the expected and observed information matrices, respectively, which are presented in Ferrari & Pinheiro (2012, p. 584). Furthermore, let  $J_{\psi\psi}$  denote the  $s \times s$  observed information matrix corresponding to  $\psi$ . Similarly,  $A_{\psi\psi}$  denotes a matrix formed from the  $(r + s) \times (r + s)$  matrix  $A$  by dropping the rows and columns that correspond to the parameter of interest. Additionally, hat and tilde symbols indicate evaluation at the unrestricted ( $\hat{\theta}$ ) and at the restricted ( $\tilde{\theta}$ ) maximum likelihood estimator of  $\theta$  for a given  $\psi$ , respectively. For instance,  $\hat{I} = I(\hat{\theta})$ ,  $\tilde{I} = I(\tilde{\theta})$ , and  $\hat{J} = J(\hat{\theta})$ .

Inference on  $\nu$  can be based on the signed likelihood ratio statistic

$$R = R(\nu) = \text{sgn}(\hat{\nu} - \nu) \{2[\ell(\hat{\theta}) - \ell(\tilde{\theta})]\}^{1/2}.$$

It is well known that  $R$  has the first order of accuracy only, i.e.  $R$  is approximately distributed as a standard normal distribution with error of order  $O(n^{-1/2})$ . This approximation can be inaccurate when the sample size is small. The accuracy of this approximation can be improved by making suitable adjustments to  $R$ .

Barndorff-Nielsen (1986, 1991) proposed a modified signed likelihood ratio statistic,  $R^*$ , which has an asymptotic normal distribution with approximation error of order  $O(n^{-3/2})$ . Significance tests and confidence limits based on the normal approximation to the distribution of  $R^*$  are, in general, extremely accurate even for very small sample sizes. The proposed modification is defined in terms of derivatives of the log-likelihood function with respect to the sample space, more precisely, derivatives with respect to the maximum likelihood estimator of  $\theta$ . To obtain the derivatives, one needs to specify an ancillary statistic, for example  $\mathbf{a}$ , such that  $(\hat{\theta}, \mathbf{a})$  is a sufficient statistic. Thus, the log-likelihood function can be written as  $\ell(\theta; \hat{\theta}, \mathbf{a})$  and one can derive  $\ell(\theta; \hat{\theta}, \mathbf{a})$  with respect to  $\hat{\theta}$ . In the following, we use the standard notation for derivatives of the log-likelihood function:  $\ell_\theta(\theta) = \ell_\theta(\theta; \hat{\theta}, \mathbf{a}) = \partial \ell(\theta; \hat{\theta}, \mathbf{a}) / \partial \theta^\top$ , and  $\ell_{;\hat{\theta}}(\theta) = \ell_{;\hat{\theta}}(\theta; \hat{\theta}, \mathbf{a}) = \partial \ell(\theta; \hat{\theta}, \mathbf{a}) / \partial \hat{\theta}$ . Hence,  $\ell_\theta(\theta)$  is a  $(k+m)$  column vector,  $\ell_{;\hat{\theta}}(\theta)$  is a  $(k+m)$  row vector, and  $\ell_{\theta;\theta}(\theta)$  and  $\ell_{\theta;\hat{\theta}}(\theta)$  are  $(k+m) \times (k+m)$  matrices.

Barndorff-Nielsen's modified signed likelihood ratio statistics is given by

$$R^* = R + \frac{1}{R} \ln \left( \left| \frac{U}{R} \right| \right), \quad (8)$$

where

$$U = \frac{\left| \begin{array}{c} \ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\hat{\theta}}(\tilde{\theta}) \\ \ell_{\psi;\hat{\theta}}(\tilde{\theta}) \end{array} \right|}{|\tilde{J}_{\psi\psi}|^{1/2} |\hat{J}|^{1/2}}.$$

Here  $\ell_{\psi;\hat{\theta}}(\tilde{\theta})$  is obtained from  $\ell_{\theta;\hat{\theta}}(\tilde{\theta})$  excluding the row that corresponds to the parameter of interest  $\nu$ . For the one-parameter case, i.e. the case in which there is no nuisance parameter, we have

$$U = |\ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\hat{\theta}}(\theta)| |\hat{J}|^{-1/2}.$$

Because the computation of  $U$  involves sample space derivatives, which, in turn, requires the determination of a suitable ancillary statistic, it can be difficult or even impossible to obtain. Most of the applications of Barndorff-Nielsen's adjustment concern exponential family models and transformation models (e.g. Wong & Wu, 2002, Larsen et al., 2002, Melo & Ferrari, 2010 and Cortese & Ventura, 2013). Approximations to  $R^*$  that avoid the need for sample space derivatives have been proposed by several authors. The different approximations lead to alternative adjusted signed likelihood ratio statistics. It is our aim to obtain and compare the different adjusted statistics for the extreme value regression models introduced in Section 2.

DiCiccio & Martin (1993) proposed an approximation to Barndorff-Nielsen's modified statistic that requires that the parameter of interest and the nuisance parameters be globally orthogonal, i.e., the Fisher information matrix is block-diagonal. It should be noted that it

is always possible to obtain an orthogonal parameterization when the parameter of interest is scalar (Cox & Reid, 1987). The adjusted signed likelihood ratio statistic proposed by DiCiccio & Martin (1993), denoted here by  $R_0^*$ , is defined by (8) with  $U$  replaced by

$$U_0 = \ell_\nu(\tilde{\theta}) \left( \frac{|\tilde{J}_{\psi\psi}| |\hat{I}_{\nu\nu}|}{|\hat{J}| |\tilde{I}_{\nu\nu}|} \right)^{1/2}, \quad (9)$$

the sign of  $U_0$  being taken to be the same as that of  $R$ , where  $\ell_\nu(\theta)$  is the derivative of the log-likelihood function with respect to the parameter of interest. According to Severini (2000, Section 7.5.2),  $R_0^* = R^* + O_p(n^{-1})$  for  $\nu$  of the form  $\hat{\nu} = \nu + O(n^{-1})$ .

The adjusted signed likelihood ratio statistic proposed by Skovgaard (1996), denoted here by  $\bar{R}^*$ , is based on covariances and is given by (8) with  $U$  replaced by

$$\bar{U} = \left| \begin{pmatrix} q^\top \\ \Upsilon_\psi \end{pmatrix} \right| |\hat{I}|^{-1} |\hat{J}|^{1/2} |\tilde{J}_{\psi\psi}|^{-1/2}, \quad (10)$$

where  $q = q(\hat{\omega}, \tilde{\omega})$  and  $\Upsilon = \Upsilon(\hat{\omega}, \tilde{\omega})$ , with

$$q(\omega_1, \omega) = E_{\omega_1}[U(\omega_1) (\ell(\omega_1) - \ell(\omega))], \quad \Upsilon(\omega_1, \omega) = E_{\omega_1}[U(\omega_1) U^\top(\omega)].$$

Also,  $\Upsilon_\psi$  is the matrix  $\Upsilon$  without the row that corresponds to the parameter of interest  $\nu$ . According to Severini (2000, Section 7.5.4),  $\bar{R}^* = R^* + O_p(n^{-1})$  for  $\nu$  of the form  $\hat{\nu} = \nu + O(n^{-1})$ .

The adjusted signed likelihood ratio statistic proposed by Severini (1999), denoted here by  $\hat{R}^*$ , is based on empirical covariances and is given by (8) with  $U$  replaced by

$$\hat{U} = \left| \begin{pmatrix} \hat{q}^\top \\ \hat{\Upsilon}_\psi \end{pmatrix} \right| |\hat{I}|^{-1} |\hat{J}|^{1/2} |\tilde{J}_{\psi\psi}|^{-1/2}, \quad (11)$$

where  $\hat{q}$  and  $\hat{\Upsilon}$  are empirical estimates of  $q$  and  $\Upsilon$  given by

$$\hat{q} = \sum_{t=1}^n (\ell_t(\hat{\theta}) - \ell_t(\tilde{\theta})) \frac{\partial \ell_t(\hat{\theta})}{\partial \theta^\top} \quad \text{and} \quad \hat{\Upsilon} = \sum_{t=1}^n \frac{\partial \ell_t(\tilde{\theta})}{\partial \theta} \frac{\partial \ell_t(\hat{\theta})}{\partial \theta^\top}. \quad (12)$$

According to Severini (2000, Section 7.5.5),  $\hat{R}^* = R^* + O_p(n^{-1})$  for  $\nu$  of the form  $\hat{\nu} = \nu + O(n^{-1})$ .

The adjusted signed likelihood ratio statistic given by Fraser, Reid & Wu (Fraser et al., 1999), denoted here by  $\tilde{R}^*$ , is based on an approximate ancillary statistic, and is defined by (8) with  $U$  replaced by

$$\tilde{U} = \frac{\left| \begin{array}{c} \tilde{\ell}_{;\hat{\theta}}(\hat{\theta}) - \tilde{\ell}_{;\hat{\theta}}(\tilde{\theta}) \\ \tilde{\ell}_{\psi;\hat{\theta}}(\tilde{\theta}) \end{array} \right|}{|\tilde{J}_{\psi\psi}|^{1/2} |\hat{J}|^{1/2}},$$

where

$$\tilde{\ell}_{;\hat{\theta}}(\hat{\theta}) - \tilde{\ell}_{;\hat{\theta}}(\tilde{\theta}) = \left( \ell_{;y}(\hat{\theta}) - \ell_{;y}(\tilde{\theta}) \right) \hat{V} \left( \ell_{\theta;y}(\hat{\theta}) \hat{V} \right)^{-1} \hat{J}$$

is an  $1 \times (k + m)$  vector,

$$\tilde{\ell}_{\theta;\hat{\theta}}(\tilde{\theta}) = \ell_{\theta;y}(\tilde{\theta})\hat{V} \left( \ell_{\theta;y}(\hat{\theta})\hat{V} \right)^{-1} \hat{J}$$

is a  $(k + m) \times (k + m)$  matrix,  $\ell_{;y}$  is the derivative of the log-likelihood function with respect to  $y$ ,  $\ell_{\theta;y}$  is its second derivative with respect to  $\theta$  and  $y$ , and  $\hat{V}$  is an  $n \times (k + m)$  matrix with the element  $(i, j)$  given by  $(-\partial F(y_i, \theta)/\partial \theta_j) / f(y_i, \theta) \big|_{\theta=\hat{\theta}}$ ,  $F$  is the cumulative distribution function of the model and  $f$  is the corresponding probability density function. According to Severini (2000, Section 7.5.3),  $\tilde{R}^* = R^* + O_p(n^{-3/2})$  for  $\nu$  of the form  $\hat{\nu} = \nu + O(n^{-1})$ . It should be noted that the approximation error for  $\tilde{R}^*$  is smaller than those corresponding to  $R_0^*$ ,  $\bar{R}^*$  and  $\hat{R}^*$ , which are all  $O_p(n^{-1})$ .

## 4 Adjusted signed likelihood ratio statistics for extreme value regression models

First, we note that the adjusted signed likelihood ratio statistics presented in the previous section require the computation of the observed and/or expected information matrices. Both matrices for extreme value regression models can be found in Ferrari & Pinheiro (2012).

It is always possible to obtain Barndorff-Nielsen's modified statistic  $R^*$  in homoskedastic linear regression models by considering the ancillary statistic  $\mathbf{a} = (a_1, \dots, a_n)$ , where  $a_t = (y_t - x_t^\top \hat{\beta})/\hat{\sigma}$ . The linear maximum extreme value regression model with constant dispersion is a special case of model (4)-(6), in where  $g$  and  $h$  are the identity function,  $\eta_t = x_t^\top \beta$  and  $\delta_t = \sigma$ . For this model,

$$\begin{aligned} \ell_{;\hat{\theta}}(\hat{\theta}) - \ell_{;\tilde{\theta}}(\tilde{\theta}) &= \iota^\top \left( \hat{\sigma}^{-1}(-\mathcal{I} + \hat{\mathcal{Z}}) - \tilde{\sigma}^{-1}(-\mathcal{I} + \tilde{\mathcal{Z}}) \right) \begin{pmatrix} \hat{X} & \hat{\mathcal{Z}}_\iota \end{pmatrix}, \\ \ell_{\theta;\hat{\theta}}(\tilde{\theta}) &= \frac{\partial^2 \ell(\theta; \hat{\theta}, a)}{\partial \theta \partial \hat{\theta}} \bigg|_{\tilde{\theta}} = \tilde{A} \begin{pmatrix} \hat{X} & \hat{\mathcal{Z}}_\iota \end{pmatrix}, \end{aligned}$$

where

$$A = \sigma^{-2} \begin{pmatrix} X^\top \check{\mathcal{Z}} \\ \iota^\top (\mathcal{I} - \check{\mathcal{Z}} + \mathcal{Z} \check{\mathcal{Z}}) \end{pmatrix},$$

$X = (x_1^\top, \dots, x_n^\top)^\top$ ,  $\iota$  is an  $n$ -dimensional column vector of ones,  $\mathcal{I}$  is the  $n \times n$  identity matrix,  $\mathcal{Z} = \text{diag}\{z_1, \dots, z_n\}$ ,  $\check{\mathcal{Z}} = \text{diag}\{\exp(-z_1), \dots, \exp(-z_n)\}$  and  $z_t = (y_t - \mu_t)/\sigma_t$ . Hence, Barndorff-Nielsen's modified statistic  $R^*$  can be written as in (8) with

$$U = \left| \begin{pmatrix} \iota^\top \left( \hat{\sigma}^{-1}(-\mathcal{I} + \hat{\mathcal{Z}}) - \tilde{\sigma}^{-1}(-\mathcal{I} + \tilde{\mathcal{Z}}) \right) \\ \tilde{A}_\psi \end{pmatrix} \begin{pmatrix} \hat{X} & \hat{\mathcal{Z}}_\iota \end{pmatrix} \right| |\tilde{J}_{\psi\psi}|^{-1/2} |\hat{J}|^{-1/2},$$

where  $\tilde{A}_\psi$  is obtained from  $A$  by excluding the row that corresponds to  $\nu$ .

We now obtain the aforementioned approximations to  $R^*$  for the maximum extreme value regression model (4)-(6). It should be noted that, hereafter the linearity and constant dispersion assumptions are no longer required.

One should take the following steps to obtain DiCiccio & Martin's adjusted signed likelihood ratio statistic  $R_0^*$  (DiCiccio & Martin, 1993): find a parameterization for which the parameter of interest is orthogonal to the remaining parameters; re-write (4)-(6) in the orthogonal parameterization; obtain the log-likelihood function, its derivative with respect to the parameter of interest, the observed information matrix and the element of the expected information matrix that corresponds to the parameter of interest; compute (9). It should be noted that the adjustment term depends on the choice of the parameter of interest and the systematic sub-models. In the Appendix we present the steps above for inference on one of the components of the  $\beta$  vector in a model with linear specification for the location and dispersion sub-models.

Skovgaard's adjusted signed likelihood ratio statistic  $\bar{R}^*$  (Skovgaard, 1996) is obtained by replacing  $q$  and  $\Upsilon$  in (10) with

$$\bar{q} = \begin{bmatrix} \hat{X}^\top \hat{\Phi}^{-1} \hat{T} C (\mathcal{I} - M \check{D}) \iota \\ \hat{Z}^\top \hat{\Phi}^{-1} \hat{H} (C(\mathcal{E}\mathcal{I} + N \check{D}) - \mathcal{I}) \iota \end{bmatrix}$$

and

$$\bar{\Upsilon} = \begin{bmatrix} \hat{X}^\top \hat{\Phi}^{-1} \hat{T} C M \check{D} \tilde{T} \tilde{\Phi}^{-1} \tilde{X} & \hat{X}^\top \hat{\Phi}^{-1} \hat{T} C \{\mathcal{I} + \check{D}(-M - CN + MD)\} \tilde{H} \tilde{\Phi}^{-1} \tilde{Z} \\ -\hat{Z}^\top \hat{\Phi}^{-1} \hat{H} C N \check{D} \tilde{T} \tilde{\Phi}^{-1} \tilde{X} & \hat{Z}^\top \hat{\Phi}^{-1} \hat{H} C \{\mathcal{E}\mathcal{I} + \check{D}(N + CP - ND)\} \tilde{H} \tilde{\Phi}^{-1} \tilde{Z} \end{bmatrix},$$

respectively, where  $T = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$ ,  $H = \text{diag}\{1/h'(\sigma_1), \dots, 1/h'(\sigma_n)\}$ ,  $\Phi = \text{diag}\{\sigma_1, \dots, \sigma_n\}$ ,  $C = \text{diag}\{\hat{\sigma}_1/\tilde{\sigma}_1, \dots, \hat{\sigma}_n/\tilde{\sigma}_n\}$ ,  $D = \text{diag}\{(\hat{\mu}_1 - \tilde{\mu}_1)/\tilde{\sigma}_1, \dots, (\hat{\mu}_n - \tilde{\mu}_n)/\tilde{\sigma}_n\}$ ,  $\check{D} = \text{diag}\{\exp(-(\hat{\mu}_1 - \tilde{\mu}_1)/\tilde{\sigma}_1), \dots, \exp(-(\hat{\mu}_n - \tilde{\mu}_n)/\tilde{\sigma}_n)\}$ ,  $M = \text{diag}\{\Gamma(1 + \hat{\sigma}_1/\tilde{\sigma}_1), \dots, \Gamma(1 + \hat{\sigma}_n/\tilde{\sigma}_n)\}$ ,  $N = \text{diag}\{\Gamma^{(1)}(1 + \hat{\sigma}_1/\tilde{\sigma}_1), \dots, \Gamma^{(1)}(1 + \hat{\sigma}_n/\tilde{\sigma}_n)\}$ ,  $P = \text{diag}\{\Gamma^{(2)}(1 + \hat{\sigma}_1/\tilde{\sigma}_1), \dots, \Gamma^{(2)}(1 + \hat{\sigma}_n/\tilde{\sigma}_n)\}$ , and the other matrices are as defined above. Here,  $\Gamma(\cdot)$  is the gamma function, and  $\Gamma^{(1)}(\cdot)$  and  $\Gamma^{(2)}(\cdot)$  denote the first and second derivatives of the function, respectively.

To obtain Severini's adjusted signed likelihood ratio statistic  $\hat{R}^*$  (Severini, 1999),  $\hat{q}$  and  $\hat{\Upsilon}$  are replaced in (11) by

$$\hat{q} = \begin{pmatrix} \hat{X} \hat{\Phi}^{-1} \hat{T} (\hat{\mathcal{L}} - \tilde{\mathcal{L}}) (\mathcal{I} - \hat{\mathcal{Z}}) \iota \\ \hat{Z} \hat{\Phi}^{-1} \hat{H} (\hat{\mathcal{L}} - \tilde{\mathcal{L}}) (-\mathcal{I} + \hat{\mathcal{Z}} - \hat{\mathcal{Z}} \hat{\mathcal{Z}}) \iota \end{pmatrix} \quad \text{and} \quad \hat{\Upsilon} = \begin{bmatrix} \hat{\Upsilon}_{\beta\beta} & \hat{\Upsilon}_{\beta\gamma} \\ \hat{\Upsilon}_{\gamma\beta} & \hat{\Upsilon}_{\gamma\gamma} \end{bmatrix},$$

where

$$\begin{aligned} \hat{\Upsilon}_{\beta\beta} &= \tilde{X}^\top \tilde{\Phi}^{-1} \tilde{T} (\mathcal{I} - \tilde{\mathcal{Z}}) (\mathcal{I} - \hat{\mathcal{Z}}) \hat{\Phi}^{-1} \hat{T} \hat{X}, \\ \hat{\Upsilon}_{\beta\gamma} &= \tilde{X}^\top \tilde{\Phi}^{-1} \tilde{T} (\mathcal{I} - \tilde{\mathcal{Z}}) (-\mathcal{I} + \hat{\mathcal{Z}} - \hat{\mathcal{Z}} \hat{\mathcal{Z}}) \hat{\Phi}^{-1} \hat{H} \hat{Z}, \\ \hat{\Upsilon}_{\gamma\beta} &= \tilde{Z}^\top \tilde{H} \tilde{\Phi}^{-1} (-\mathcal{I} + \tilde{\mathcal{Z}} - \tilde{\mathcal{Z}} \tilde{\mathcal{Z}}) (\mathcal{I} - \hat{\mathcal{Z}}) \hat{\Phi}^{-1} \hat{T} \hat{X}, \end{aligned}$$



$$\hat{\Upsilon}_{\gamma\gamma} = \tilde{Z}\tilde{H}\tilde{\Phi}^{-1}(-\mathcal{I} + \tilde{Z} - \tilde{Z}\tilde{Z})(-\mathcal{I} + \hat{Z} - \hat{Z}\hat{Z})\hat{\Phi}^{-1}\hat{H}\hat{Z},$$

$\mathfrak{L} = \text{diag}\{\ell_1, \dots, \ell_n\}$ , with

$$\ell_t = \ell_t(\mu_t, \sigma_t) = \ln f(y_t; \mu_t, \sigma_t) = -\ln \sigma_t - (y_t - \mu_t)/(\sigma_t) - \exp(-(y_t - \mu_t)/(\sigma_t)),$$

and the other matrices are as defined above.

Finally, Fraser, Reid & Wu's adjusted signed likelihood ratio statistic  $\tilde{R}^*$  (Fraser et al., 1999) is given by (8) with  $U$  replaced by

$$\tilde{U} = \left| \begin{pmatrix} \iota \left( \hat{\Phi}^{-1}(-\mathcal{I} + \hat{Z}) - \tilde{\Phi}^{-1}(-\mathcal{I} + \tilde{Z}) \right) \\ \tilde{A}_{\psi} \end{pmatrix} \begin{pmatrix} \hat{T}\hat{X} & \hat{Z}\hat{H}\hat{Z} \end{pmatrix} \begin{pmatrix} \hat{A}(\hat{T}\hat{X} & \hat{Z}\hat{H}\hat{Z}) \end{pmatrix}^{-1} \right| |\tilde{J}_{\psi\psi}|^{-1/2} |\hat{J}|^{1/2},$$

where

$$A = \begin{pmatrix} X^{\top} T \Phi^{-2} \check{Z} \\ Z^{\top} H \Phi^{-2} (\mathcal{I} - \check{Z} + Z\check{Z}) \end{pmatrix},$$

$\hat{A} = A(\hat{\theta})$ ,  $\tilde{A} = A(\tilde{\theta})$ ,  $\tilde{\ell}_{\psi; \tilde{\theta}}(\tilde{\theta})$  is obtained from  $\tilde{\ell}_{\theta; \tilde{\theta}}(\tilde{\theta})$  by excluding the row corresponding the parameter of interest  $\nu$ , and the other matrices are as defined above.

All of the formulas presented above are valid for maximum extreme value regression models. If, instead, the data are assumed to follow a minimum extreme value regression model, the formulas can be easily adapted. Recall that, from (3), if  $y_t \sim EV_{\min}(\mu_t, \sigma_t)$ , then  $y_t^* = -y_t$  has a maximum extreme value distribution with parameters  $-\mu_t$  and  $\sigma_t$  ( $-y_t \sim VE_{\max}(-\mu_t, \sigma_t)$ ). Hence, one can obtain all of the adjusted statistics by taking the response variable  $y_t^*$ , and the systematic components  $g^*(\mu_t) = \mu_t^* = \eta_t^* = -g^{-1}(\eta(x_t, \beta))$  and  $h^*(\sigma_t) = h(\sigma_t) = \delta(z_t, \gamma)$ . In other words, adjusted signed likelihood ratio statistics for the minimum extreme value regression model (7) with systematic components (5) and (6) can be computed from the results above by changing the signs of the observations on the response variable and using an identity link function for the location parameter with the modified predictor  $\eta_t^*$ . We note that Fraser, Reid & Wu's adjusted statistic for the minimum extreme value distribution (iid case) under censoring was obtained by Wong & Wu (2000).

## 5 Monte Carlo simulation results

We now present a Monte Carlo simulation study to evaluate the finite-sample performance of the signed likelihood ratio test ( $R$ ), Barndorff-Nielsen's modified version ( $R^*$ ) and the tests that use approximations to  $R^*$ , namely DiCiccio & Martin's  $R_0^*$ , Skovgaard's  $\bar{R}^*$ , Severini's  $\hat{R}^*$  and Fraser, Reid & Wu's  $\tilde{R}^*$ , in extreme value regression models. The maximum likelihood estimates of the parameters are obtained by numerically maximizing the log-likelihood function using the nonlinear quasi-Newton BFGS method with analytical derivatives (see, for instance, Press et al., 1992), implemented in the function MaxBFGS in the matrix programming language

0x (Doornik, 2009). The number of Monte Carlo replicates is 10,000 and the nominal levels of the tests are  $\alpha = 10\%$ ,  $5\%$  and  $1\%$ .

At the outset, we consider the maximum extreme value regression model with constant dispersion and location sub-model

$$\mu_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4};$$

‘model 1’. The hypothesis being tested is  $\mathcal{H}_0 : \beta_1 \leq 0$  and the alternative hypothesis is  $\mathcal{H}_1 : \beta_1 > 0$ . We set  $\beta_0 = 1$ ,  $\beta_1 = 0$ ,  $\beta_2 = 1$ ,  $\beta_3 = 6$ ,  $\beta_4 = -3$  and  $\sigma = 1$ . The values of the covariates are randomly drawn from the  $\mathcal{U}(-0.5, 0.5)$  distribution. The sample sizes range from 15 to 200.

Table 1 presents the null rejection rates of the six tests. The figures in this table indicate that the signed likelihood ratio test is liberal, more so when the sample size is small. It is noteworthy that all of the adjusted tests are much less size distorted. For instance, when  $n = 20$  and  $\alpha = 5\%$ , the null rejection rate is  $8.1\%$  whereas the rejection rates of the adjusted tests are  $5.0\%$  (Barndorff-Nielsen’s  $R^*$ , DiCiccio & Martin’s  $R_0^*$  and Fraser, Reid & Wu’s  $\tilde{R}^*$ ),  $4.0\%$  (Skovgaard’s  $\bar{R}^*$ ) and  $3.8\%$  (Severini’s  $\hat{R}^*$ ). As expected, the null rejection rates of all the tests are near the nominal levels for large samples.

Figure 1 presents plots of relative p-value discrepancies versus asymptotic p-values for small to moderate sample sizes, namely  $n = 15, 20, 30$ . The relative p-value discrepancy is the difference between the exact and the asymptotic p-values divided by the latter. It should be noted that, on the left-hand side of the plots (asymptotic p-value smaller than  $10\%$ ) the relative discrepancy line of the signed likelihood ratio test is well above the zero line. Hence, the plots show that the signed likelihood ratio test is anti-conservative, as anticipated by the figures presented in Table 1. The adjusted tests are clearly better behaved, particularly Barndorff-Nielsen’s  $R^*$ , DiCiccio & Martin’s  $R_0^*$  and Fraser, Reid and Wu’s  $\tilde{R}^*$ . Their relative p-value discrepancies are not far from zero, even for very small asymptotic p-values (e.g.,  $0.1\%$ ) and small sample size (e.g.,  $n = 15$ ). The plots also show that the other tests, Skovgaard’s  $\bar{R}^*$  and Severini’s  $\hat{R}^*$ , are conservative when  $n = 15$  and  $20$ . The conservative behavior is more pronounced for  $\bar{R}^*$ . When  $n = 30$  the p-value discrepancies for both tests are very small.

We now turn to power simulations. As noted above, the different tests have different sizes when the asymptotic normal distribution is used as an approximation to the true null distribution of the statistics. To guarantee that all tests have the correct size, we first simulate 500,000 samples under the null hypothesis being tested. We then obtain the empirical distribution of each statistic under the null hypothesis, and we estimate the exact critical values of tests that give the correct size for the chosen nominal level. The exact critical values are used to evaluate the power of the tests. This strategy is used for all of the power simulations in this paper. Table 2 presents the rejection rates of the tests under the alternative hypotheses  $\beta_2 = \epsilon$  for different values of  $\epsilon > 0$ . It is apparent that all of the tests have very similar powers.

Table 1: Null rejection rates (%), model 1

$n$	$\alpha = 10\%$						$\alpha = 5\%$						$\alpha = 1\%$					
	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R^*$
15	13.8	10.0	9.0	6.8	10.0	10.0	8.4	4.9	4.2	3.3	4.9	4.9	2.7	1.0	0.7	0.6	1.0	1.0
20	13.5	9.9	8.8	8.3	9.9	9.9	8.1	5.0	4.0	3.8	5.0	5.0	2.3	1.1	0.8	0.7	1.1	1.1
30	13.7	10.8	10.1	10.1	10.8	10.8	7.7	5.2	4.8	4.8	5.2	5.2	2.2	1.2	0.9	1.0	1.2	1.2
40	12.2	9.6	8.8	9.7	9.6	9.6	6.7	4.8	4.3	4.8	4.8	4.8	1.7	1.2	1.0	1.1	1.2	1.2
100	10.5	9.6	9.2	9.5	9.6	9.6	5.2	4.7	4.6	4.8	4.7	4.7	1.2	0.9	0.8	0.9	0.9	0.9
200	10.5	10.1	10.0	10.2	10.1	10.1	5.3	5.0	4.8	5.0	5.0	5.0	1.2	1.0	0.9	1.1	1.0	1.0

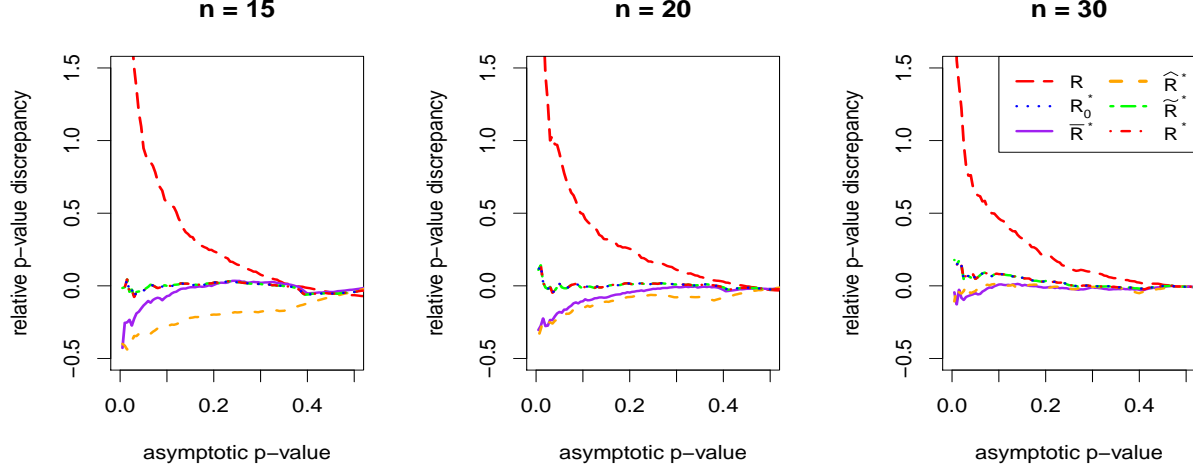


Figure 1: Relative p-value discrepancy plots, model 1.

Table 2: Non-null rejection rates (%), model 1,  $n = 20$ ,  $\alpha = 10\%$ 

$\epsilon$	0.1	0.5	1.0	2.0	3.0
$R$	11.7	22.3	39.8	76.5	94.6
$R_0^*$	11.7	22.2	39.8	76.9	94.8
$\bar{R}^*$	11.7	22.1	39.8	76.8	94.8
$\hat{R}^*$	11.7	21.6	37.9	72.5	91.8
$\tilde{R}^*$	11.7	22.2	39.8	76.9	94.8
$R^*$	11.7	22.2	39.8	76.9	94.8

We now consider the maximum extreme value regression model with location and dispersion sub-models

$$\mu_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3}$$

and

$$\ln(\sigma_t) = \gamma_0 + \gamma_1 z_{t1} + \gamma_2 z_{t2};$$

‘model 2’. The null and alternative hypotheses are  $\mathcal{H}_0 : \beta_3 \leq 0$  and  $\mathcal{H}_1 : \beta_3 > 0$ . Here,

we set  $\beta_0 = 1$ ,  $\beta_1 = 1$ ,  $\beta_2 = 6$ ,  $\beta_3 = 0$ ,  $\gamma_0 = 1$  and  $\gamma_1 = \gamma_2 = 0.1$ , where the values of the covariates are drawn from a  $\mathcal{U}(-0.5, 0.5)$  distribution, and the sample sizes are 40, 50, 60 and 70. For this model and for model 3 (below), it is not possible to obtain the adjusted signed likelihood ratio statistic  $R^*$ . Therefore, we only consider the approximations to  $R^*$  presented in the previous section. The size and power simulation results are summarized in Tables 3 and 4 and in Figure 2. Again, it is clear that the signed likelihood ratio test is oversized when the sample is not large, the proposed adjustments are effective in pushing the size to the chosen nominal level, all of the rejection rates converge to the nominal level as  $n$  grows, and the tests are equally powerful.

Table 3: Null rejection rates (%), model 2

$n$	$\alpha = 10\%$					$\alpha = 5\%$					$\alpha = 1\%$				
	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$
40	12.0	10.8	10.2	9.0	9.5	6.9	5.8	5.5	4.3	4.9	1.9	1.4	1.3	0.8	1.0
50	11.7	10.6	9.6	9.3	9.5	6.4	5.7	4.9	4.7	4.8	1.6	1.3	0.9	0.8	0.9
60	11.6	10.9	10.1	9.5	9.9	6.3	5.7	4.9	4.7	5.0	1.3	1.1	0.9	0.7	0.8
70	11.4	11.0	10.2	9.7	10.2	5.9	5.6	4.9	4.7	4.9	1.4	1.3	1.1	0.9	1.0
100	10.7	10.3	9.7	9.7	9.7	5.9	5.6	5.0	5.1	5.2	1.2	1.1	0.9	0.9	1.0
200	10.8	10.7	10.2	10.4	10.4	5.4	5.3	5.0	5.0	5.0	1.1	1.1	1.0	1.0	1.0

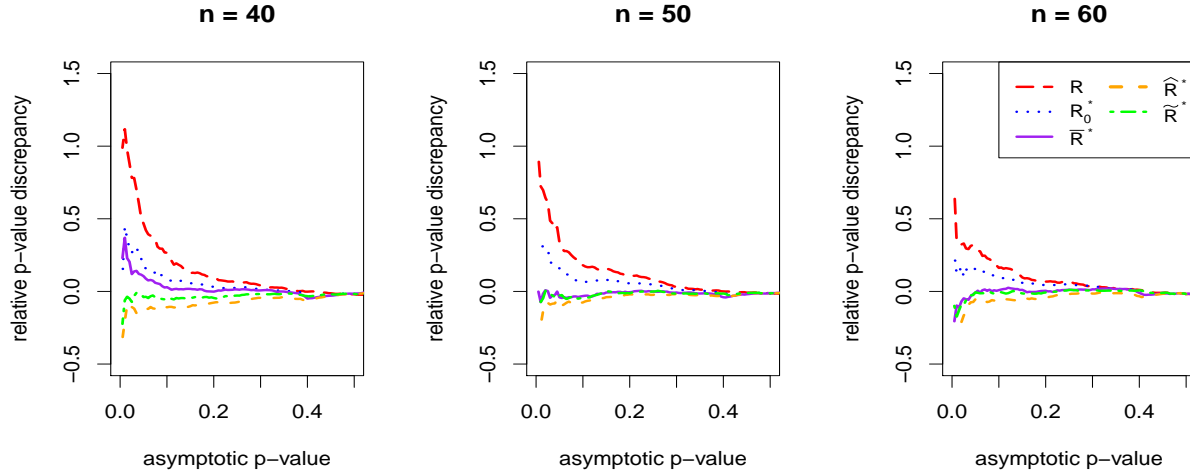


Figure 2: Relative p-value discrepancy plots, model 2.

We now consider the maximum extreme value regression model with constant dispersion and a nonlinear specification for the location parameter given by

$$\mu_t = \beta_0 + \beta_1 x_{t1} + x_{t2}^{\beta_2};$$

Table 4: Non-null rejection rates (%), model 1,  $n = 40$ ,  $\alpha = 10\%$ 

$\epsilon$	0.1	0.5	1.0	2.0	4.0	6.0
$R$	10.4	14.4	21.5	39.9	77.6	95.2
$R_0^*$	10.3	14.7	21.9	40.5	78.2	95.8
$\bar{R}^*$	10.3	14.6	21.7	40.0	78.3	95.8
$\hat{R}^*$	10.4	14.4	21.2	39.2	76.2	94.5
$\tilde{R}^*$	10.4	14.7	21.6	40.3	78.3	95.6

‘model 3’. The null hypothesis is  $\mathcal{H}_0 : \beta_2 \leq 0$  and the alternative hypothesis is one-sided. We set  $\sigma = 1$ ,  $\beta_0 = 1$ ,  $\beta_1 = 1$  and  $\beta_2 = 0$ , where the covariate values are randomly drawn from a  $\mathcal{U}(0, 1)$  distribution, and the sample sizes are 15, 20, 30 and 40. The simulation results are presented in Tables 5 and 6 and in Figure 3. As in the previous scenarios, the liberal behavior of the signed likelihood ratio test, the effectiveness of the adjustments and the similar power behavior of the different tests are evident. It should be noted, however, that the tests that use Severini’s  $\hat{R}^*$  and Fraser, Reid and Wu’s  $\tilde{R}^*$  statistics perform notably better than the other tests when the sample is small.

Table 5: Null rejection rates (%); model 3

$n$	$\alpha = 10\%$					$\alpha = 5\%$					$\alpha = 1\%$				
	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$
15	14.2	12.0	8.7	10.6	10.8	8.1	6.6	3.9	5.1	5.5	2.2	1.7	0.7	1.1	1.1
20	13.3	10.7	9.0	10.3	10.5	7.2	5.3	4.3	5.0	5.1	2.1	1.3	0.9	1.2	1.2
30	12.1	10.4	9.4	10.5	10.4	6.8	5.3	4.6	5.3	5.2	1.8	1.2	1.0	1.2	1.2
40	11.7	10.3	9.7	10.7	10.2	6.5	5.3	4.7	5.5	5.2	1.6	1.1	1.0	1.2	1.1
100	10.5	10.0	9.7	10.0	10.0	5.9	5.4	5.1	5.4	5.4	1.2	1.0	0.9	1.1	1.0

Table 6: Non-null rejection rates (%); model 3,  $n = 15$ ,  $\alpha = 10\%$ 

$\epsilon$	0.01	0.03	0.05	0.10	0.20	0.30
$R$	12.0	16.2	21.1	37.5	76.5	96.3
$R_0^*$	12.0	16.3	20.9	37.3	76.6	96.4
$\bar{R}^*$	12.0	16.0	20.7	37.1	76.1	96.5
$\hat{R}^*$	11.8	16.3	21.0	36.8	75.5	95.6
$\tilde{R}^*$	12.0	16.2	20.9	37.3	76.5	96.3

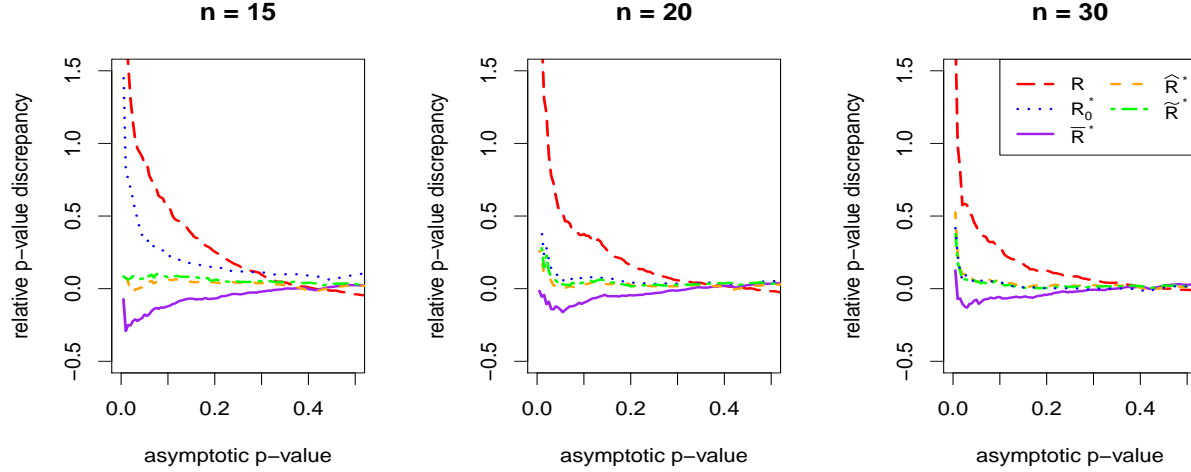


Figure 3: Relative p-value discrepancy plots, model 3.

## 6 Applications

Our first application involves with a data set (see Table 7), consisting of maximum wind speed measured in January from 2001 to 2010 and the minimum temperature on the day in which the maximum wind speed was reached. The data were extracted from the alpine tundra climate station (3743 m) of Niwot Ridge/Green Lakes Valley, Colorado, USA.<sup>1</sup>

Table 7: Maximum wind speed (m/s) and minimum temperature (°C) in January

year	temperature	wind speed
2001	−7.40	33.42
2002	−11.95	44.04
2003	−17.99	42.92
2004	−25.63	42.51
2005	−16.61	45.75
2006	−10.93	47.78
2007	−9.21	43.34
2008	−26.13	48.69
2009	−20.27	43.20
2010	−19.00	43.00

We assume that the maximum wind speed follows a maximum extreme value regression

<sup>1</sup>CR23X data: <http://niwot.colorado.edu/exec/.extracttoolA?d-1cr23x.ml>.

model (4) with constant dispersion and location sub-model given by

$$\mu_t = \beta_0 + \beta_1 x_t,$$

for  $t = 1, \dots, 10$ , where the covariate  $x$  is the minimum temperature.

Table 8 presents the maximum likelihood estimates for the parameters and the corresponding standard errors. The values of the test statistics for testing the null hypothesis  $\mathcal{H}_0 : \beta_1 \geq 0$  against  $\mathcal{H}_1 : \beta_1 \leq 0$  and the corresponding p-values are presented in Table 9. We note that the p-value of the unmodified signed likelihood ratio test is 1.1% while the p-values of the modified tests range from 2.8% (Fraser, Reid and Wu's  $\tilde{R}^*$  and Barndorff-Nielsen's  $R^*$ ) to 5.4% (Skovgaard's  $\bar{R}^*$ ). The unmodified test displays the smallest p-value, in accordance with its liberal behavior observed in our simulation study. The different adjustments weaken the evidence in favor of the alternative hypothesis.

Table 8: Estimates and standard errors, wind speed data set

Parameters	$\beta_0$	$\beta_1$	$\sigma$
	intercept	minimum	
	temperature		
estimate	34.3412	-0.4409	3.4211
s.e.	3.0910	0.1740	0.8435

Table 9: Test statistics, , wind speed data set

	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R^*$
statistic	-2.2912	-1.8989	-1.6085	-1.7592	-1.9043	-1.9043
p-value	0.0110	0.0288	0.0539	0.0393	0.0284	0.0284

Our second application involves with a data set consisting of 34 men's decathlon results at the 1988 Olympic Games gathered by Hand (1996, p. 304, data set 357).<sup>2</sup> We assume that the score in high jump follows a maximum extreme value regression model (4) with constant dispersion and location sub-model

$$\mu_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \beta_5 x_{t5},$$

for  $t = 1, \dots, 34$ . The covariates are the scores in the following events: javelin throw ( $x_1$ ), long jump ( $x_2$ ), discus throw ( $x_3$ ), shot put ( $x_4$ ), and pole vault ( $x_5$ ).

The maximum likelihood estimates for the parameters and the corresponding standard errors are presented in Table 10, and the values of test statistics for testing the null hypotheses

<sup>2</sup>The data set is available at [http://www.stat.ncsu.edu/working\\_groups/sas/sic1/data/olympic.dat](http://www.stat.ncsu.edu/working_groups/sas/sic1/data/olympic.dat).

$\mathcal{H}_0 : \beta_1 \leq 0$  and  $\mathcal{H}_0 : \beta_4 \geq 0$  are presented in Table 11. For both cases, the p-values of the signed likelihood ratio test is much smaller than the p-values of the adjusted tests. It is interesting to observe that conclusions based on the unadjusted test may differ from those based on adjusted versions. For instance, at the 5% nominal level, the signed likelihood ratio test ( $R$ ) rejects  $\mathcal{H}_0 : \beta_1 \leq 0$ , whereas all of the adjusted tests do not reject it. In addition, at the 10% nominal level, only the unadjusted test rejects the null hypothesis  $\mathcal{H}_0 : \beta_4 \geq 0$ . Based on our simulation results it is advisable to rely on the adjusted tests.

Table 10: Estimates and standard errors, decathlon data set

Parameters	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\sigma$
	intercept	javelin	long	discuss	shot	pole	
		throw	jump	throw	put	vault	
estimate	-0.1971	0.2125	0.0522	0.0371	-0.1844	0.1522	0.0383
s.e.	0.3873	0.0889	0.1674	0.1244	0.1291	0.0951	0.0051

Table 11: Test statistics, decathlon data set

	$R$	$R_0^*$	$\bar{R}^*$	$\hat{R}^*$	$\tilde{R}^*$	$R^*$
$\mathcal{H}_0 : \beta_1 \leq 0$ versus $\mathcal{H}_1 : \beta_1 > 0$						
statistic	2.0102	1.5841	1.5658	1.5920	1.5945	1.5945
p-value	0.0222	0.0566	0.0587	0.0557	0.0554	0.0554
$\mathcal{H}_0 : \beta_4 \geq 0$ versus $\mathcal{H}_1 : \beta_4 < 0$						
statistic	-1.4815	-1.1743	-1.1277	-0.5062	-1.1789	-1.1789
p-value	0.0692	0.1201	0.1297	0.3064	0.1192	0.1192

## 7 Conclusion

In this paper we derived Barndorff-Nielsen's modified signed likelihood ratio statistic for the homoskedastic linear extreme value regression model, and different approximations to this statistic in the general case were made, i.e., possibly nonlinear and/or heteroskedastic extreme value regression models. Unlike the other adjusted statistics, DiCiccio & Martin's statistic requires an orthogonal parameterization, which makes its deduction more involved. Our simulation results revealed that the signed likelihood ratio test is liberal in small and moderate-sized samples, i.e., it leads to a type I error probability greater than the chosen nominal level. Additionally,



it is evident from the simulations that all of the adjustments considered in this paper are effective in shrinking the size distortion of the original test. For the linear homoskedastic model, Barndorff-Nielsen's, DiCiccio's and Fraser, Reid & Wu's tests behaved equally well and clearly better than Severini's and Skovgaard's tests. For the nonlinear homoskedastic model, Severini's and Fraser, Reid & Wu's tests presented better performance than the others. In terms of power, all of the tests behaved similarly. Overall, Fraser, Reid & Wu's test is the best performing test.

Our applications show that the conclusion reached by the signed likelihood ratio test may conflict with the conclusion achieved by the adjusted tests when the sample is not large (in our application,  $n = 10$  and  $34$ ). As indicated by our simulations, practitioners should rely on the adjusted tests.

## Acknowledgments

We gratefully acknowledge the financial support from the Brazilian agencies CNPq, CAPES and FAPESP.

## Appendix

Let  $v$  be a vector, and  $v_{(r)}$  be the vector  $v$  without its  $r$ th component. Analogously, let  $\mathcal{V}$  be a matrix and  $\mathcal{V}_r$  be the matrix  $\mathcal{V}$  without the  $r$ th column. Assume that  $\nu = \beta_r$  is the parameter of interest. The Fisher information matrix (see [Ferrari & Pinheiro, 2012](#)) can be written as

$$I(\theta) = I(\beta_r, \theta_{(r)}) = I(\beta_r, \beta_{(r)}, \gamma) = \begin{pmatrix} I_{\beta_r \beta_r} & I_{\beta_r \beta_{(r)}} & I_{\beta_r \gamma} \\ I_{\beta_{(r)} \beta_r} & I_{\beta_{(r)} \beta_{(r)}} & I_{\beta_{(r)} \gamma} \\ I_{\gamma \beta_r} & I_{\gamma \beta_{(r)}} & I_{\gamma \gamma} \end{pmatrix},$$

where

$$\begin{aligned} I_{\beta_r \beta_r} &= \mathbf{x}_{\cdot r}^\top \Phi^{-1} T^2 \Phi^{-1} \mathbf{x}_{\cdot r}, & I_{\beta_r \beta_{(r)}} &= I_{\beta_{(r)} \beta_r}^\top = \mathbf{x}_{\cdot r}^\top \Phi^{-1} T^2 \Phi^{-1} X_{(r)}, \\ I_{\beta_r \gamma} &= I_{\gamma \beta_r}^\top = (\mathcal{E} - 1) \mathbf{x}_{\cdot r}^\top \Phi^{-1} T H \Phi^{-1} Z, & I_{\beta_{(r)} \beta_{(r)}} &= X_{(r)}^\top \Phi^{-1} T^2 \Phi^{-1} X_{(r)}, \\ I_{\beta_{(r)} \gamma} &= I_{\gamma \beta_{(r)}}^\top = (\mathcal{E} - 1) X_{(r)}^\top \Phi^{-1} T H \Phi^{-1} Z, & I_{\gamma \gamma} &= (1 + \Gamma^{(2)}(2)) Z^\top \Phi^{-1} H^2 \Phi^{-1} Z, \end{aligned}$$

$\mathbf{x}_{\cdot r} = (x_{1r}, \dots, x_{nr})^\top$  is the  $r$ th column of  $X$  and  $\Phi$ ,  $T$ ,  $H$ ,  $X$  and  $Z$  as defined above.

Consider a parameterization  $\vartheta = (\beta_r, \vartheta_{(r)})$ ,  $\vartheta_{(r)} = (\kappa, \tau)$ , where  $\vartheta$ ,  $\vartheta_{(r)}$ ,  $\kappa$  and  $\tau$  have dimensions  $(k + m)$ ,  $(k + m - 1)$ ,  $(k - 1)$  and  $m$ , respectively, in such a way that  $\beta_r$  is orthogonal to  $\vartheta_{(r)}$ . Let

$$I_{\beta_r \theta_{(r)}} = \begin{pmatrix} I_{\beta_r \beta_{(r)}} & I_{\beta_r \gamma} \end{pmatrix}, \quad I_{\theta_{(r)} \theta_{(r)}} = \begin{pmatrix} I_{\beta_{(r)} \beta_{(r)}} & I_{\beta_{(r)} \gamma} \\ I_{\gamma \beta_{(r)}} & I_{\gamma \gamma} \end{pmatrix}, \quad I_{\theta_{(r)} \theta_{(r)}}^{-1} = \begin{pmatrix} I^{\beta_{(r)} \beta_{(r)}} & I^{\beta_{(r)} \gamma} \\ I^{\gamma \beta_{(r)}} & I^{\gamma \gamma} \end{pmatrix}.$$

Define  $A \equiv I_{\beta_r \beta_{(r)}} I^{\beta_{(r)} \beta_{(r)}} + I_{\beta_r \gamma} I^{\gamma \beta_{(r)}}$ ,  $B \equiv I_{\beta_r \beta_{(r)}} I^{\beta_{(r)} \gamma} + I_{\beta_r \gamma} I^{\gamma \gamma}$ ,  $\beta_{(r)} = \kappa - \beta_r A^\top$  and  $\gamma = \tau - \beta_r B^\top$ . From [Cox & Reid \(1987\)](#), it can be shown that  $\beta_r$  is orthogonal to  $(\kappa, \tau)$ .

In order to obtain  $R_0^*$ , one should first re-write (4)-(6) in the orthogonal parameterization  $\vartheta = (\beta_r, \kappa, \tau)$ , and define the systematic components  $\eta^*(\vartheta)$  and  $\delta^*(\vartheta)$ . The derivative of the log-likelihood function with respect to the parameter of interest  $\beta_r$ ,  $\ell_{\beta_r}^*(\vartheta)$ , the observed information matrix,  $J^*(\vartheta)$  and the element of the expected information matrix that corresponds to  $\beta_r$ ,  $I_{\beta_r\beta_r}^*(\vartheta)$ , should then be computed in this parameterization as in Ferrari & Pinheiro (2012, p. 584), and inserted into (9).

Consider, for instance, the maximum extreme value model regression model (4) with location and dispersion sub-models  $g(\mu_t) = \eta_t = \eta(x_t, \beta) = x_t^\top \beta$  and  $h(\sigma_t) = \delta_t = \delta(z_t, \gamma) = z_t^\top \gamma$ . We have

$$g(\mu_t) = x_{tr}\beta_r + x_{t(r)}^\top \beta_{(r)} = x_{tr}\beta_r + x_{t(r)}^\top (\kappa - \beta_r A^\top) = (x_{tr} - x_{t(r)}^\top A^\top) \beta_r + x_{t(r)}^\top \kappa \equiv \eta_t^*(\vartheta) \quad (13)$$

and

$$h(\sigma_t) = z_t^\top \tau - \beta_r z_t^\top B^\top \equiv \delta_t^*(\vartheta). \quad (14)$$

For the reparameterized model (4) with (13)-(14), we have

$$\ell_{\beta_r}^*(\vartheta) = \frac{\partial \ell^*(\vartheta)}{\partial \beta_r} = \iota^\top \left( \Phi^{-1} T \left( \mathcal{I} - \check{Z} \right) \mathbf{v}_1 + \Phi^{-1} H \left( -\mathcal{I} + \mathcal{Z} - \mathcal{Z} \check{Z} \right) \mathbf{v}_2 \right),$$

$$\begin{aligned} J_{\beta_r\beta_r}^* &= \mathbf{v}_1^\top V_{\beta\beta} \mathbf{v}_1 + 2\mathbf{v}_1^\top V_{\beta\gamma} \mathbf{v}_2 + \mathbf{v}_2^\top V_{\gamma\gamma} \mathbf{v}_2, & J_{\kappa\kappa}^* &= X_{(r)}^\top V_{\beta\beta} X_{(r)}, \\ J_{\beta_r\kappa}^* &= J_{\kappa\beta_r}^\top = \mathbf{v}_1^\top V_{\beta\beta} X_{(r)} + \mathbf{v}_2^\top V_{\beta\gamma} X_{(r)}, & J_{\kappa\tau}^* &= J_{\tau\kappa}^\top = X_{(r)}^\top V_{\beta\gamma} Z, \\ J_{\beta_r\tau}^* &= J_{\tau\beta_r}^\top = \mathbf{v}_1^\top V_{\beta\gamma} Z + \mathbf{v}_2^\top V_{\gamma\gamma} Z, & J_{\tau\tau}^* &= Z^\top V_{\gamma\gamma} Z \end{aligned}$$

and

$$I_{\beta_r\beta_r}^* = \mathbf{v}_1^\top \Phi^{-1} T^2 \Phi^{-1} \mathbf{v}_1 + 2(\mathcal{E} - 1) \mathbf{v}_1^\top \Phi^{-1} T H \Phi^{-1} \mathbf{v}_2 + (1 + \Gamma^{(2)}(2)) \mathbf{v}_2^\top \Phi^{-1} H^2 \Phi^{-1} \mathbf{v}_2,$$

where  $\mathbf{v}_1 = (\mathbf{x}_r - X_{(r)} A^\top)$ ,  $\mathbf{v}_2 = -(Z B^\top)$ ,  $V_{\beta\beta} = \Phi^{-1} T \left( \check{Z} \Phi^{-1} + (\mathcal{I} - \check{Z}) S T \right) T$ ,  $V_{\beta\gamma} = \Phi^{-1} T (\mathcal{I} - \check{Z} + \mathcal{Z} \check{Z}) H \Phi^{-1}$  and  $V_{\gamma\gamma} = \Phi^{-1} H \left( (-\mathcal{I} + 2\mathcal{Z} - 2\mathcal{Z} \check{Z} + \mathcal{Z}^2 \check{Z}) \Phi^{-1} + (-\mathcal{I} + \mathcal{Z} - \mathcal{Z} \check{Z}) Q H \right) H$ . DiCiccio & Martin's adjusted signed likelihood ratio statistic can now be obtained by replacing the formulas above in (9).

## References

- Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters, based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307-322. [2](#), [3](#), [5](#)
- Barndorff-Nielsen, O.E. (1991). Modified signed log likelihood ratio. *Biometrika*, **78**, 557-563. [2](#), [3](#), [5](#)
- Barreto-Souza, W. & Vasconcellos, K.L.P. (2011). Bias and skewness in a general extreme value regression model. *Computational Statistics & Data Analysis* **55**, 3, 1379-1393. [2](#)
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. UK Chichester: John Wiley & Sons. [2](#)
- Castillo, E., Hadi, A. S., Balakrishnan, N. & Sarabia, J. M. (2005). *Extreme value and related models with applications in engineering and science*. New Jersey: John Wiley & Sons. [2](#), [3](#)

- Coles, S. (2001). *An Introduction to Statistical Modeling of Extremes*. London: Springer-Verlag. [2](#), [3](#)
- Cortese, G. & Ventura, L. (2013). Accurate higher-order likelihood inference on  $P(Y < X)$ . *Computational Statistics*, **28**, 1035-1059. [5](#)
- Cox, D. R., & Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*, **40**, 1-39. [6](#), [17](#)
- DiCiccio, T. J. & Martin, M. A., (1993). Simple modifications for signed roots of likelihood ratio statistics. *Journal of the Royal Statistical Society B*, **55**, 305–316. [2](#), [3](#), [5](#), [6](#), [8](#)
- Doornik, J. A. (2009). *An Object-Oriented Matrix Language – Ox 6*. London: Timberlake Consultants Press. [10](#)
- Dunn, P. K. & Smyth, G. (1996). Randomized quantile residuals. *Journal of the Computational and Graphical Statistics*, **5**, 236-244.
- Ferrari, S.L.P. & Pinheiro, E.C. (2011). Improved likelihood inference in beta regression. *Journal of Statistical Computation and Simulation*, **81**, 4 , 431-443.
- Ferrari, S.L.P. & Pinheiro, E.C. (2012). Small-sample likelihood inference in extreme-value regression models. *Journal of Statistical Computation and Simulation*, **84**, 3 , 582-595. [2](#), [4](#), [7](#), [17](#), [18](#)
- Fraser, D.A.S., Reid, N. & Wu, J. (1999). A simple general formula for tail probabilities for frequentist and bayesian inference. *Biometrika*, **86**, 249-264. [2](#), [3](#), [6](#), [9](#)
- Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math Cracovie*, **6**, 32, 93-116. [1](#)
- Fisher, R.A., & Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *In Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, 2, 180-190. Cambridge University Press. [1](#)
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, **44**, 423-453. *Breakthroughs in Statistics*, Vol.I (1992), eds. S. Kotz and N.L. Johnson, Springer-Verlag, pp. 195-225. [2](#), [3](#)
- Gumbel, E.J. (1958). *Statistics of Extremes*. 2nd ed. New York: Columbia University Press. [1](#)
- de Haan, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematical Centre Tracts 32. Amsterdam: Mathematics Centre. [2](#)

- Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. & Ostrowsky, E. (1996) *A Handbook of Small Data Sets*. 2nd ed. London: Chapman & Hall. 15
- Kotz, S., Nadajarah, S. (2000). *Extreme Value Distributions: Theory and Applications*. London: Imperial College Press. 2
- Larsen, P.V., Blæsild, P., & Sørensen, M.K. (2002). Improved likelihood ratio tests on the von Mises-Fisher distribution. *Biometrika*, **89**, 947-951. 5
- Lawless, J.F. (2003). *Statistical Models and Methods of Lifetime Data*. 2nd ed. New Jersey: John Wiley & Sons. 3
- Lemonte, A.J. & Ferrari, S.L.P. (2011). Signed likelihood ratio tests in the Birnbaum-Saunders regression model. *Journal of Statistical Planning and Inference*, **141**, 1031-1040.
- Melo, T.F., & Ferrari, S.L.P. (2010). A modified signed likelihood ratio test in elliptical structural models. *Advances in Statistical Analysis*, **94**, 1, 75-87. 5
- Mueller, S. & Rufibach, K. (2009). Smooth tail-index estimation. *Journal of Statistical Computation and Simulation*, **79**, 1155-1167. 2
- Press, W.H., Teulosky, S.A., Vetterling, W.T. & Flannery, B.P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. London: Prentice Hall . 9
- Severini, T.A. (1999). An empirical adjustment to the likelihood ratio statistic. *Biometrika*, **86**, 235-247. 2, 3, 6, 8
- Severini, T.A. (2000). *Likelihood Methods in Statistics*. Oxford: Oxford University Press. 6, 7
- Skovgaard, I.M. (1996). An explicit large-deviance approximation to one-parameter tests. *Bernoulli*, **2**, 145-165. 2, 3, 6, 8
- Skovgaard, I.M. (2001). Likelihood asymptotics. *Scandinavian Journal of Statistics*, **28**, 3-32. 2
- Wong, A.C.M. & Wong and Wu, J. (2000). Practical small-sample asymptotics for distributions used in life-data analysis. *Technometrics*, **42**, 149-155. 9
- Wong, A.C.M. & Wu, J. (2002). Small sample asymptotic inference for the coefficient of variation: normal and nonnormal models. *Journal of Statistical Planning and Inference*, **104**, 73-82. 5